

Enhanced Automatic 2D-3D Conversion using Retinex in Machine Learning Framework

José L. Herrera, Carlos R. del-Blanco and Narciso García

Abstract—In this paper, we present an approach for automatically convert images from 2D to 3D. The algorithm uses a color + depth dataset to estimate a depth map of a query color image by searching structurally similar images in the dataset and fusing them. Our experimental results indicate that the inclusion of a retinex based stage for the query image and the dataset images improves the performance of the system on commonly-used databases and for different image descriptors.

Keywords—2D-to-3D image conversion; depth maps; retinex;

I. INTRODUCTION

In the last years, the amount of devices with 3D play capability, such as smart phones, TVs, DVD/Blu-Ray players, video game consoles or cinemas, has experimented a significantly increase. Nevertheless, the availability of 3D content such as pictures, movies, or broadcastings has not experienced the same trend. To reduce this gap between 3D players and 3D content, different algorithms have been developed for automatically convert 2D content into 3D [1].

Recently, new machine learning-based algorithms have appeared to perform the 2D-to-3D conversion task. These new approaches are based on the core assumption that photometrically similar image have likely analogous 3D structure (depth map). In this family of conversion algorithms, the most remarkable ones are: the work of Karsch et al. [2], which employs SIFT-flow to find the most similar images, and then performs an optimization process to refine the result; the approach of Konrad et al. [3], which uses a HOG-based descriptor instead of SIFT-flow, and also performs a Cross Bilateral Filtering to enhance the edges of the result; and the works of Herrera et al [4], which extended the previous approach to efficiently use large databases through a hierarchical search, and also improved the performance of the algorithm by using an adaptive number of images in the fusion process of the depth maps [5].

A common limitation of all the previous approaches is the strong dependence that has the image descriptors (HOG, SIFT or LBP) with the illumination conditions of the acquired scenes. This fact produces that structurally similar images, but acquired with very different illumination conditions, render dissimilar feature descriptors, and thus decreasing the quality of the estimated depth map.

In this paper an automatic 2D-to-3D conversion method based on a learning approach is proposed, which solves the aforementioned problem using the retinex algorithm [6], a

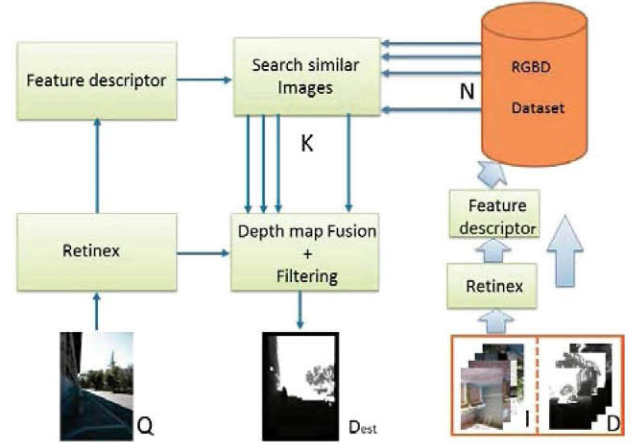


Fig. 1. Block diagram of our 2D-to-3D conversion approach

color constancy method that ensures that the perceived color remains relatively constant under varying illumination conditions. The 2D-to-3D algorithm evaluates the similarity between a query color image and the database images, combining the most similar ones to estimate the depth map of the scene. In this context, retinex algorithm is computed both for query image and for images in the database.

II. ALGORITHM DESCRIPTION

Given a query color image Q and a database $DDBB$, composed by pairs of color images + associated depth maps, the purpose is to estimate a depth map for Q . The presented algorithm, described in Fig. 1 can be split into five steps.

The first step is to compensate the illumination variations of the color images (query image and all the database images) to compute reliable and repeatable feature descriptors from them. This is accomplished by the multi-scale retinex algorithm [6], a variation of retinex that simultaneously provides local lightness/contrast enhancement, dynamic range compression, and good color rendition. Notice that this is an off-line processing for the database images, and therefore the computational performance of the whole 2D-3D conversion system is not greatly affected

The second step of the algorithm is the computation of the feature descriptors of the query image and the images in the database. As previous step, this computation may be done offline for all database images. Different descriptors have



Fig. 2. From left to right: Ground truth, Query Image, and Depth Estimation.

been evaluated: a HOG-based descriptor (as in Konrad’s work [3]), and a LBP-based descriptor (as in Herrera’s [4] method).

In the third step, the similarity between the descriptor from the query image and the ones from the database images is computed, using the correlation as similarity metric. The depth maps associated to the K most similar images (the ones with highest correlation coefficients) are selected for the next step.

In the fourth step, the K selected depth maps, which theoretically are the most structurally similar, are fused using a weighted average. The weights are obtained from the computed correlation scores in the third step.

Finally, in the fifth step, the resulting depth map fusion is enhanced by a Cross Bilateral Filtering that highlights the edges, and align them with the edges of the query image.

III. EXPERIMENTAL RESULTS

The proposed approach has been tested using the Make3D dataset [7], which consists in a set of 400 training images and 134 test images; and the NYU dataset [8], composed by 1449 images. To evaluate the quality of the generated depth maps, the normalized cross covariance (C), the $PSNR$, and the Structural Similarity ($SSIM$) have been adopted as metrics. In all the metrics, the highest value represents the highest quality.

Fig. 2 shows two examples of the results of the conversion process performed by our approach. Tables 1 show the results of the algorithm over the Make3D dataset, using the training and test subsets independently and in a Leave-One-Out (LOO) configuration, and over the NYU dataset in LOO configuration for the three quality metrics. From the results, we can conclude that the correction of the illumination conditions via retinex improves the quality of the estimated depth map.

IV. CONCLUSIONS

An automatic method for 2D-to-3D image conversion from a single color image based on machine learning has been presented. A module of illumination compensation has been

TABLE I. QUALITY METRICS FOR THE DIFFERENT DATABASES

Make3D dataset (134 test, 400 train)[7]			
Algorithm	C	$PSNR$	$SSIM$
HOG	0.647	13.379	0.742
HOG_retinex	0.652	13.862	0.773
LBP	0.665	14.181	0.779
LBP_retinex	0.683	14.381	0.780
Make3D dataset (Leave One Out)[7]			
HOG	0.619	13.178	0.722
HOG_retinex	0.626	14.034	0.763
LBP	0.638	14.491	0.775
LBP_retinex	0.650	14.575	0.775
NYU (Leave One Out)[8]			
HOG	0.561	12.951	0.789
HOG_retinex	0.606	13.536	0.801
LBP	0.606	13.560	0.803
LBP_retinex	0.608	13.574	0.803

included using the retinex theory. Two different feature descriptors (HOG-based and LBP-based) have been tested, and for both cases, the inclusion of the retinex module improves the performance of the algorithm. This improvement is obtained because the compensation of the illumination in the scenes renders similar image feature descriptors for those images that have similar 3D structure, although they have been acquired with very different illumination conditions. This fact affects to the selection of the K most similar images, resulting in a better estimation of the depth of the query image.

REFERENCES

- [1] C. Cheng, C. Li, and L. Chen, "A novel 2D-to-3D conversion system using edge information," IEEE Trans. On Consumer Electronics, vol.56, no.3, pp.1739,1745, Aug. 2010.
- [2] K. Karsch, C. Liu, and S. B. Kang, "Depth Transfer: Depth Extraction from Video Using Non-Parametric Sampling," IEEE Trans.on Pattern Anal. and Mach. Intell., vol.36, no.11, pp.2144-58, Nov. 1 2014.
- [3] J. Konrad, M. Wang, P. Ishwar, C. Wu, and D. Mukherjee, "Learning-based, automatic 2d-to-3d image and video conversion," IEEE Trans. on Image Process., vol. 22, no. 9, pp. 3485–3496, Sept 2013.
- [4] J.L. Herrera, C.R. del Blanco, and N. Garcia, "Learning 3d structure from 2d images using lbp features," in Int. Conf. on Image Process. October 2014, pp. 2022–2025.
- [5] J.L. Herrera, C.R. del Blanco, and N. Garcia, "Fast 2d to 3d conversion using a clustering-based hierarchical search in a machine learning framework," in IEEE 3DTV Conference, 2014, July 2014, pp. 1–4.
- [6] B. Jiang, G. A. Woodell, and D. J. Jobson, "Novel multi-scale retinex with color restoration on graphics processing unit," Journal of Real-Time Image Processing, pp. 1-15, 2014.
- [7] A. Saxena, M. Sun, and A.Y. Ng, "Make3d: Learning 3d scene structure from a single still image," IEEE Trans.on Pattern Anal. and Mach. Intell., vol. 31, no. 5, pp.824–840, May 2009
- [8] N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor," IEEE Inter. Conf in Computer Vision, Nov 2011, pp. 601–608.